

# Egy “nyelvészbarát” szövegfeldolgozó eszköz: a NooJ

*Pajzs Júlia*

## **Bevezetés**

A NooJ nyelvészek számára készült, szövegfeldolgozásra alkalmas szoftver, amelynek használatával nyelvtechnológiai háttérrel nem rendelkező bölcészek is önállóan felépíthetik és akár tartalmi szempontból is elemezhetik saját korpuszaikat. A korpuszokon kívül a felhasználók maguk készíthetik el saját szótáraikat, grammatikai szabályrendszerüket is, a NooJ környezet lehetővé teszi az újonnan készített források rendszerbe illesztését. A felhasználás megkönnyítése érdekében már számos nyelvre, így a magyarra is elkészültek az alapvető források, morfológiai információt tartalmazó szótárak és egyszerű nyelvtanok, amelyek a szoftverhez hasonlóan kutatási célból szabadon letölthetők és felhasználhatók. A szoftver lehetővé teszi, hogy ne csupán szavakat vagy szókombinációkat, hanem bizonyos általunk definiált szintaktikai szerkezeteket vagy akár szemantikai és szintaktikai jellemzők által együttesen meghatározott példákat kereshessünk a feldolgozandó szövegekben (pl. kommunikációval vagy érzelemkifejezéssel kapcsolatos ígétet, ember által készített dolgok tárgyragos alakjait, minőségjelzős szerkezetek tartalmazó mondatokat). Nyelvészeti képzésben is hasznos lehet a NooJ: a szintaktikai gráfok készítése és az ezek segítségével talált példák vizsgálata elősegítheti a nyelvész hallgatók formalizált gondolkodásra való képességének fejlesztését.

## **2. Előzmények**

A NooJ rendszer elődjét különféle nyelvfeldolgozási feladatok megoldására fejlesztették az 1990-es években a Maurice Gross vezette University Paris 7 Laboratoires d’Automatique Documentaire et Linguistique munkatársai. Az első változat neve INTEX volt, amelyet később két irányban fejlesztettek tovább. A Sébastien Paumier és Eric Laporte által továbbfejlesztett változat az UNITEX (<http://igm.univ-mlv.fr/~unitex/>), a Max Silberstein által továbbfejlesztett pedig a NooJ ([www.nooj-association.org](http://www.nooj-association.org)), amelyhez Váradi Tamás vezetésével az MTA NYTI Nyelvtechnológia Kutatócsoportjának munkatársai készítettek magyar szótárakat (<http://corpus.nytud.hu/nooj/corpus.htm>).

## **3. Általános ismertetés**

A NooJ legalapvetőbb funkciója, hogy folyó szövegekből gyakorisági listát és konkordanciát készít. Ha van az illető nyelvhez olyan NooJ formátumú szótár, amely morfológiai információkat tartalmaz, akkor képes az ún. ‘Linguistic Analysis’ funkció végrehajtására (1. ábra). Ez alapesetben annyit jelent, hogy felismeri a szövegszavakat, a szöveg-szó szótövét, és azonosítja az aktuális toldalékokat. A magyar szótárakban lévő morfológiai kódrendszer Elekfi László – Laczkó Krisztina (2006) munkáján alapul.

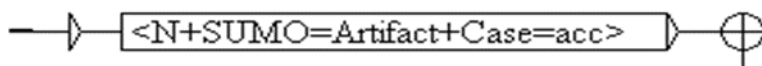
Language is "Hungarian(hu)".  
Text Delimiter is: \n (NEWLINE)  
Text contains 6004 Text Units (TUs).  
78244 tokens including:

varázsolni, cipőnkrol egy újságpapirossal leszedtünk egy kevés agyagot, és vártuk Zolit. De Zoli nem jött. A kedves osztrák barátunk is visszajött értünk, türelmetlenül topogott. Zoli nem jött.  
- Majd én bemegyek érte - mondta a kedves Heinrich, és megzörgette a férfimosdó ajtaját.  
És két perc múlva valóban megjelent Zoli.  
Ugyanolyan koszosan, ugyanolyan csuromvizesen, gyűrött ingben... kezében az óriási herendi vázával.  
- A ruháimat tettem le - mondta elcsukló hangon.  
- Barátain, barátaim! - sürgetett minket Heinrich.  
- Most már mindegy. Gyerünk.

57	63	72	75
kevés, NUM	agyag, N+DOMAIN=building_industry+SUMO=Clay+Number=sg+Case=acc	és, CON	vár, V+

**1. ábra: Részlet a Text-> Linguistic Analysis művelet eredményéből: a kevés agyagot, és szövegrészlet elemezve**

Szintaktikai és szemantikai jellegű szabályok is megfogalmazhatók. Minden szabály többféle formában is leírható: gráfként vagy „szöveges” formában, reguláris kifejezések segítségével (2. ábra). (A gráfokban használt jeleket lásd a mellékletben.)



**2. ábra: Egyszerű gráffal történő lekérdezés: ember által készített dolgok tárgyragos alakjai**

A szabályokban megfogalmazott tételekre illeszkedő szövegrészeket kikeresi, és konkordanciaként megjeleníti a találatokat. A szabályra illeszkedő nyelvi egységekkel műveleteket is végezhetünk: megjelölhetjük például a felismert főnévi csoport határát vagy a tagmondat, mondat határát (3. ábra).

felforrt a hűtővíz, egyszer kellett	kereket	ragasztani, néhányszor leállt a motor
töltötte el amúgy is csüggedő	szívünket	. Mentünk, zötyögtünk az ismeretlen úton
Nem emlékszem, hogy raktuk ki	csomagjainkat	. Egy ruhásbőrönd. Rajta egy másik
is, már csak kis fehérneműs	csomagjainkat	cipeitük, és ázott ballonkabátban, sártól
én nem mocskolom be a	palcámat	. Kilépnél! (Az öreg előrelép) És
dolgozom? Mit áll? Hozza a	csomagomat	az autóbushoz! PADRONE Uram, órákig
adni, mint a kőrmöm feketéje. (	Cédulát	néz.) A svájci kiadónál annyit
meg a nagybátyámat... ezt a	csomagot	cipelem neki fél Európán keresztül
már megjelent. De maga tud	figurákat	teremteni... egy ilyen süket öregúr
ital, fáradt vagyok. COLLE Takarót,	párnát	talál a szekrény aljában. Nem
szerettek egy bűnöst? Kivágnék egy pár	figurát	kartonból. Dehogy vacakolok vele. Itt
én félek... félek. (Felveszi a	csomagját	és a kabátját, óvatosan átlép
neki adni a horgolt ruhás	babát	, ami ott ül a dívány
rögtön végigterült a vászonhuzatos fotelon,	lábát	egy székre rakta, és az
csak a rekamiét és a	polcot	kell megtartani, no meg a
tátogó szájnak a főzelékkel teli	kanalat	. Forró nyári nap van. Verácska
a birsalmaszagot, a kopott, barna	bútorokat	. És utálta a nyálás puszkát
az eleje hibás, annak a	hátát	ragasztanám össze egy olyan Valamivel
békésen megegyezni, hogy ezeket a	bútorokat	kivághassuk innen. Ide jön majd
minden fiú először a lányok	lábát	nézi? Sebő, ne egyen annyi
és gyilkos feketekávé nem támadná	szívét	, tüdejét. Egy kert, persze, egy
miatt a járdán valaki a	lábát	töri, azért büntetőjogilag felelős. A
Minek? - Javítani akarok rajta egy	sort	. Jogom... - Majd kapsz levonatot. - Most
(az író barátja, aki négy	sort	el akart helyezni róla az
takarékoskodtak, hónapokon át nézték a	kirakatokat	, tervezgettek, latolgattak, rajzolgattak, kimérték az
-bajjal fel akarta vinni a	csomagokat	a vonatra, akkor a mögötte
Paul levelét mutatták. Yvette barátságos	sorait	, az unokák fényképét, az útlevelet
amink még volt, kulacsot, magyar	babát	, hogy bevigyük mentőangyalunknak. Az irodát

### 3. ábra: Ember által készített dolgok tárgyrágos alakjainak konkordanciája

Nagy Viktor (Vajda – Nagy – Dancsecs 2004) készítette el azt a rendszert, amelynek segítségével NooJ formátumú magyar szótárakat állíthatunk elő, bővíthetünk, módosíthatunk. Bár a létrehozott morfológiai szótár természetesen elvileg sem fedheti le az összes lehetséges magyar szóalakot, de nagyon sok, akár többszörösen képzett, ragokkal, jelekkel ellátott alakot felismer. Az 50 ezer névszóból kiinduló, a NooJ program segítségével előállított (kompilált) szótárváltozat például több mint 200 millió toldalékolt alak felismerésére képes. (Például felismeri a *szép* szótó *széppé*, *szépségesre*, *szépségeseket*, *szépségűek*, stb. alakjait.) A névszói szótáron kívül külön állományban találjuk az igék szótárát, ez 15 ezer alapigéből kiindulva 30 millió képzett és ragozott alakot ismer fel. E két alapszótáron kívül külön állományokban találunk nem toldalékolandó szavakat, egy másikban gyakori tulajdonneveket, egy újabbban gyakori rövidítéseket. Ha az általunk használt korpusz elemzése során gyakran ismétlődő fel nem ismert szavakat találunk – ezeket a NooJ egy Unknown nevű listába teszi, – készíthetünk további szótárt magunknak, amely csak a fel nem ismert elemeket tartalmazza a számunkra szükséges információkkal ellátva, a NooJ kézikönyvben definiált formátumban. Például néhány ismeretlen szövegszó az általam vizsgált szövegből elhelyezhető egy dic kiterjesztésű egyszerű szöveg formátumú állományban minimális információval ellátva így:

ablakfelelősnek,ablakfelelős,N+SG+DAT  
 adminisztrátorlánnyal, adminisztrátorlány,N+SG+INS  
 akácgallyaktól, akácgally,N+PL+ABL

Az így készülő szótárt felhasználás előtt a Nooj → Dictionary → Compile utasítással kell a NooJ rendszer számára használhatóvá tennünk, majd a NooJ → Info → Preferences → Lexical Analysis → Add Resources utasítás segítségével érhetjük el, hogy a további elemzéseknél a NooJ figyelembe vegye ezt is. Több, különböző méretű és jellegű korpuszt elemeztem a <http://corpus.nytud.hu/nooj/corpus.htm> honlapon elérhető szótárak segítségével, az általam tesztelt szövegekben előfordult szövegszavak 95 százalékát ismerte fel a NooJ a Lexical Analysis eredményeként. A felismerés eredménye a lexikai annotálás, például a *szépbe* szövegszónak a Lexical Analysis által előállított annotált változata:

<LU	lexikai egység eleje
LEMMA="szép"	szótó=szép
CAT="A"	szófaj=melléknév
FLX="N12B9"	morfológiai kód= N12B9
DOMAIN="quality"	szemantikai osztály=minőség
SUMO="SubjectiveAssessmentAttribute"	javasolt egyesített ontológia=szubjektív értékelés
Case="ill"	eset=illativusz
Number="sg">	szám=egyes szám
szépbe	szövegszó
</LU>	lexikai egység vége

Ahhoz, hogy a fenti annotált szóról ne pusztán morfológiai jellegű információkat (Case="ill" Number="sg") tároljon a rendszer, szemantikai információkkal is kiegészítettem a korábban Nagy Viktor és programjai segítségével előállított NooJ szótárakat. E feladathoz kézenfekvő forrásnak bizonyult a magyar WordNet adatbázis (Miháltz – Hatvani – Kuti – Szarvas – Csirik – Prószy – Váradi 2008).

#### 4. A WordNet és a szemantikus információk

A Magyar WordNetet a Miller (1995) és Fellbaum (1998) által alkotott Princeton WordNetre alapozva készítette el a Szegedi Egyetem Informatikai Tanszékcsoportja, a MTA NYTI Nyelvtechnológia Kutatócsoportja és a Morphologic Kft. egy közös projektum keretében. Maga a WordNet eredetileg a mentális lexikon számítógépes modelljeként született meg. A szavakból fogalmi hálót alkottak, amelyeknek csomópontjai szinonimahalmozok, synsetek (pl. egy synset elemei: *nagyszerű, pompás, remek, kiváló, szuper*). A különböző csomópontok között szemantikai relációk lehetnek (hyponymia?, hyponymia), és esetenként nagyobb szemantikai osztályokba sorolják őket, mint a fentebb bemutatott *szép* szó elemzett változatában látható DOMAIN="quality" 'szemantikai osztály=minőség', SUMO="SubjectiveAssessmentAttribute" 'javasolt egyesített ontológia=szubjektív értékelés' besorolások. Ha például kigyűjtjük a Magyar Wordnetből a DOMAIN=acoustics 'szemantikai osztály=akusztika'jeggyel bíró szavakat, ilyen elemeket láthatunk: *duruzsoló, morajló, susogó, nagy, mély, elordított, elkiabált, fülhasogató, fülsiketítő, fűlsüketítő, süketítő, mennydörgő hang, (emberi hang), bariton, hangtompító, halk szavú, szelíd szavú, háttérzaj, bariton, (halk, halk hangú) (nesztelen, halk járású; nesztelenül lépkedő), hangmagasság, lármás, zajos, csend, csönd, csendessé, hang,*

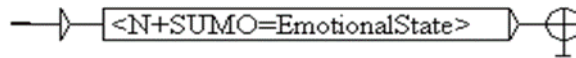
*hangoskodó, nagyhangú, nagyszájú, teleszájú hang, (cincogó, nyikorgó, csikorgó), énekhang, akusztikus.* Ebben a listában főként mellékneveket látunk, de főnév is akad a szavak között (*csend*). A WordNet-adatbázis jó alapanyag lehet egy régóta esedékes magyar fogalomköri szótárhoz, amely nem ábécé sorrendben, hanem hierarchikus struktúrába rendezve írja le a szavakat. Készíthetünk a WordNetből kiindulva ún. produktív szótárt is, amely nem ábécé sorrendben és nem hierarchikusan, hanem kulcsszavak köré rendezve írja le a hasonló jelentésű szavakat (pl. az *ad* kulcsszó alá rendezve az *adományoz, ajándékoz, átad, átenged* stb. ), esetleg szókapcsolatokat (*rendelkezésre bocsát*), közérthető definíciókkal és korpuszból származó idézetekkel segítve az adott situációban a legmegfelelőbb szó kiválasztását.

A WordNet-információkkal kiegészített NooJ szótárunkból kikereshetjük például azokat a szavakat, amelyeknek szemantikai jellemzője HYPERNYM=emotion 'hiperníma=érzelem'.

Ekkor ilyen elemeket láthatunk a szótárból:

félelem,N+FLX=N7B+HYPERNYM=emotion+NEAR\_ANTONYM=fearlessness+  
DOMAIN=factotum+SUMO=EmotionalState+ENG=fear  
öröm,N+FLX=N3C3+HYPERNYM=emotion+NEAR\_ANTONYM=sorrow+  
DOMAIN=factotum+SUMO=EmotionalState+ENG=joy.

Az így jellemzett szavak konkordancialistáját is elkészíttethetjük a NooJ segítségével, akár gráfokkal, akár reguláris kifejezésekkel definiálva lekérdezésünket. Az egyszerű gráffal való lekérdezésre és eredményére láthatunk példát a 4. és 5. ábrán. Az 5. ábra konkordancialistáján feltűnnek oda nem illő találatok is, mivel a szavak egy részének több lehetséges elemzése is van (pl. *méreg, tűz*),és a szövegek a lekérdezés előtt nem lettek egyértelműsítve.



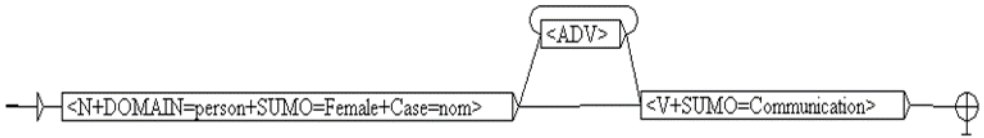
4.ábra. Gráf: érzelmi állapotra utaló főnevek lekérdezése

Istenem, milyen boldogok!	Ez van -	Örömmel	utazol? - Nem. A feleségemmel. Fénykép
munkát a vakációban?	Elpirultam az	örömtől	. Hogyne vállalnék. Boldogan. Nagyon kellett
zsebkendő. A szíve körül szűrő	fájdalom		. Egy belvárosi csemegebolt kirakatában ital
Az édesség finom. Az édesség	öröm		. Én a Gyurikámat úgy neveltem
vacak nyalánság azt a sok	fájdalmat		. - Kérlek, az én Gyurikám huszonnyolc
sárgarépa, alma, káposztasaláta. - Az étkezés	öröm		. Mindenki egye azt, ami ízlik
legyintett, mi ez a várható	gyönyörökhöz		képe? Az adásvételi szerződéssel be
szenvedek, de van értelme a	szenvedésnek		. Tél van, de jön a
filozofikus költészettel foglalkozik, némi előlegért	örömmel		írna gondolati poémákat Elidegenedő család
vásárolt egy permetezőkannát, hat kiló	mérget		, és elkecseregette szívvel pusztítani kezdte
kellene? - Az illusztráció miatt - mondta	kínban		a főszerkesztő. - Gyorsabban menne a
ifjúkori szerelmeit, vágyait és tapasztalatait,	szenvedélyeit		, életét és halálát. Tizenhét esztendeig
egy teljes estét betöltő színmű.	Szenvedélyek		. Cselekmény. Váratlan fordulatok. Happy end
fényképét, az útlevelet, és sírtak	örömlükben		. Tizennyolc éves volt Paul, amikor
itt átmennek ezen a hídon,	keresztül		a Torre di Fine-i
gyerekkocsik, biciklik. Vidám népvándorlás, integetés,	jókedv		volt az országút. És ezen
és nyálkahártyánk. - Nem... köszönjük - mondtuk	kínban		. - Előélt nem. - Valami franciás előélt
Az ember életében egyszer autózhat	keresztül		az Appennineken. Utójáték: Két nappal
Bölcskey, az ember ne legyen	haragban		a lakóbizottsággal. Leszálltunk egy bőröndöt
a millió kérdőlapot, és az	örömtől		kissé bódultan indult hazafelé. „Jütland
a millió kérdőlapot. Biztosan fehér	kereszt		van benne. Vagy kék? Ezek
északi államok mind olyan lobogót	tűznek		ki, mintha most indulnának útra
úveg is. És ekkor, keserves	kínomban		, a következőket mondtam: - Hát tudod
és mit sem törődve szepegő	félelmemmel		, anyám elővette a könyvét, apám
is tanultam, mégpedig két esztendőn	keresztül		, egy Erzsébet körüli államilag engedélyezett
játékunkat. A kicsi különben mélységes	haraggal		rugdosta a pedált, gyűlölte a
édesanyám első tekintete a rémült	csalódás		volt. Ő is, szegény, a
gyönyörű darab ementált. Lizi néni	hálából		rögtön kikelt az ágyból, levitte

## 5. ábra. Érzelmi állapotról utaló főnevek konkordanciája

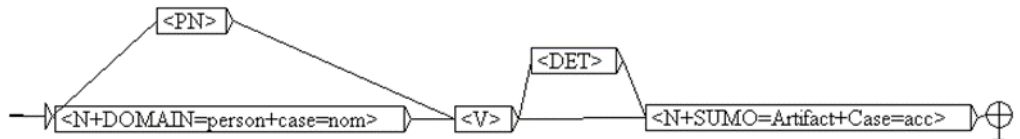
### 5. Lekérdezés NooJ gráfokkal

Nyelvészek számára főként azért lehet vonzó a NooJ program használata, mert gráfokban fogalmazhatjuk meg lekérdezéseinket, amelyek emlékeztetnek a formális grammatikákban használt nyelveírásra. A 6. ábrán szereplő gráf segítségével olyan példákat keressünk, ahol egy kommunikációval kapcsolatos igei állítmány egy nőnemű személyre referáló alanyesetben lévő főnév után áll. Az igei állítmány előtt egy vagy több határozó is lehet. Ez a leírási mód nem zárja ki, hogy a leírt elemek előtt vagy után további elemek is legyenek a mondatban (például jelző vagy számnév az alany előtt, határozó az ige után). Ezt a lekérdezést az Ittész (2006, 2011, 2014) szerkesztése alatt készülő *A magyar irodalmi és köznyelv nagyszótára* eddig megjelent kötetének példamondataiból előállított korpuszra alkalmazva, egyebek közt ilyen példamondat részleteket kaptam: *néhány tartózkodó természetű hölgy csevegett, halkan, dekoltáltan... Vera barátnője másképp beszél, ... a cseléd szóljon hozzá... mikor a mama beszél...*



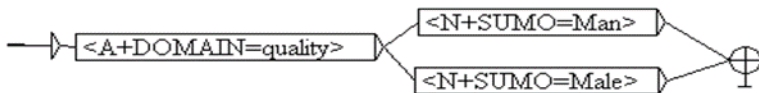
6. ábra: Gráf: nőnemű személy kommunikál.

A 7. ábra gráfja egy olyan lekérdezésre példa, ahol az alany személy (PN=PersonName=személynév), az igei állítmányt pedig egy tárgyestben lévő, mesterségesen előállított tárgyat jelölő főnév követi, amelyet megelőzhet egy névelő. Erre a lekérdezésre a Nagyszótárból ilyen példákat kaptam: *az ember lerontja az oltárokat... Rákosi vezette a labdát, Albertnek adta.*



7. ábra: Gráf: személy alanyesetben +ige+ ember által készített dolgok tárgyragos alakja

A 8. ábrán látható gráf segítségével azokat a jelzős szerkezeteket kereshetjük ki, ahol a melléknév szemantikai jegye a DOMAIN=quality, a főnévé pedig SUMO=Man, azaz hímnemű élőlényeket jelölő főnevekkel alkotott minőségjelzős szerkezeteket. Hasonló gráffal kereshetjük ki a nőnemű élőlényeket jelölő főnevekkel alkotott minőségjelzős szerkezetek is. Az eredményül kapott szókapcsolatokat összegezzük a NooJ → Concordance → Statistical Analyses utasítással, az eredményt exportálhatjuk táblázatos formába vagy egyszerű szövegfornába. Az így nyert két táblázatot összehasonlítva azt láthatjuk, hogy a Nagyszótár példamondataiban a *hölgyek* és *leányok* főként *fiatalok* és *szépek*, az *emberek* és *férfiak* leginkább *fiatalok* vagy *öreg*ek, de vannak közöttük *gyarló*k, *kedvesek*, *elevenek* és *unalmasok* is.



8. ábra. Gráf: minőségjelző hímnemű élőlényre utaló főnév előtt.

Hasonlóképpen előállíthatunk olyan táblázatot a Nagyszótár példamondataiból vagy bármilyen más korpuszból, amely az összes minőségjelzős szerkezetet tartalmazza. A kapott táblázatot akár a jelzőre, akár a jelzett szóra rendezhetjük, így hasonló adatbázist kaphatunk a tipikus jelzős szerkezetekről, mint amilyeneket az igékről a Sass és társai (2010) által készített *Magyar igei szerkezetek* tartalmazott. Például a *mély* melléknevet követő

leggyakrabban előforduló főnév az *álom, árok, csend, hang, seb, tisztelet* (1. táblázat). A főnévre rendezett táblázatból egyebek között azt láthatjuk, hogy a *munka* főnév leggyakoribb jelzője a *nagy* és a *nehéz* (2. táblázat).

1. táblázat: Részlet a minőségjelzős összetételek táblázatából a jelzőre rendezve

mély	álomba	3
mély	árokba	4
mély	csendbe	3
mély	hangok	4
mély	sebet	3
mély	tisztelettel	4

2. táblázat: Részlet a minőségjelzős összetételek táblázatából a jelzett szóra rendezve

nagy	munka	3
nehéz	munka	4
nehéz	munkát	3
nagy	munkával	4

## Összegzés

A NooJ szövegfeldolgozó program alapfunkcióinak ismertetésével a szövegfeldolgozás iránt érdeklődő kutatóknak kívántam kedvet csinálni az önálló korpuszepítéshez és -vizsgálathoz. Bár a fentiekben a magyar szövegekre való alkalmazásra mutattam csak példát, a NooJ honlapjára látogatva számos más nyelvre fejlesztett szótárakat és egyéb segédanyagokat találhatnak: magán a NooJ használatáról szóló angol nyelvű kézikönyvön kívül bőséges szemléltető anyagot és referenciát. A NooJ felhasználók évente különböző helyszíneken rendezett konferenciákon mutatják be egymásnak legfrissebb fejlesztéseiket. A NooJ programon túl a Magyar WordNet felhasználására is bátorítom a kutatókat.

\*\*\*

Életem meghatározó momentuma volt, hogy a '70-es években Bánréti Tanár Úr az akkori Pécsi Tanárképző Főiskola tanársegédjeként a Bevezetés a nyelvtudományba tantárgy és a szerda délutáni szabadon választható szűkebb körű szeminárium keretében megismertetett minket az akkori modern nyelvészettel, ezen belül figyelmembe ajánlotta az Általános Nyelvészeti Tanulmányok II. kötetét, amelynek címe: Matematikai nyelvészet. Már a címéből éreztem: ez az igazán nekem való szakterület. Ezúton szeretném megköszönni Tanár Úrnak, hogy erre a szakterületre vezetett, amelyben lehetőségem nyílt olyan feladatokkal foglalkozni, amelyeket élvezettel csinálhattam egy inspiráló környezetben.



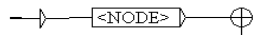
**Melléklet:** A NooJ gráfokban használt jelek magyarázata



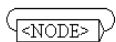
kezdőpont



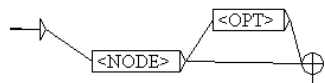
végpont



csomópont két állapot között



a csomópontban megnevezett elem többször egymás után ismétlődhet



az OPT-tal jelölt elem jelenléte opcionális

**Hivatkozások**

- Elekfi, László, Laczkó, Krisztina. 2006. Ragozási táblázatok In: Ittész Nóra. (szerk.). Nagyszótár I. Segédletek. Budapest: MTA Nyelvtudományi Intézet. 797–1114.
- Ittész, Nóra. (szerk.). 2006., 2011., 2014. Nagyszótár I.–V. Budapest: MTA Nyelvtudományi Intézet.
- Fellbaum, Christinane (ed.). 1998. WordNet: An Electronic Lexical Database Cambridge, MA: MIT Press.
- Miháltz, Márton, Hatvani, Csaba, Kuti, Judit, Szarvas, György, Csirik, János, Prószéky, Gábor, és Váradi, Tamás 2008. Methods and Results of the Hungarian WordNet Project. In: Proceedings of the Fourth Global WordNet Conference GWC 2008, pp. 310–320.
- Miller, George A. 1995. WordNet: A Lexical Database for English. In: Communications of the ACM Vol. 38, No. 11: 39–41.
- Sass, Bálint, Váradi, Tamás, Pajzs, Júlia és Kiss, Margit. 2010. Magyar igei szerkezetek. A leggyakoribb vonzatok és szókapcsolatok szótára. Tinta Könyvkiadó, Budapest.
- Vajda, Péter, Nagy Viktor és Dancsecs, Erzsébet. 2004. A Ragozási szótártól a NooJ morfológiai moduljáig. In: Alexin Zoltán – Csendes Dóra (szerk.) A II. Magyar Számítógépes Nyelvészeti Konferencia előadásai. Szeged: Szegedi Tudományegyetem. p. 183–190.